A Groupwise Multilinear Correspondence Optimization for 3D Faces

Timo Bolkart
Saarland University, Germany
tbolkart@mmci.uni-saarland.de

Stefanie Wuhrer
Inria Rhône-Alpes, France
stefanie.wuhrer@inria.fr

Abstract

Multilinear face models are widely used to model the space of human faces with expressions. For databases of 3D human faces of different identities performing multiple expressions, these statistical shape models decouple identity and expression variations. To compute a high-quality multilinear face model, the quality of the registration of the database of 3D face scans used for training is essential. Meanwhile, a multilinear face model can be used as an effective prior to register 3D face scans, which are typically noisy and incomplete. Inspired by the minimum description length approach, we propose the first method to jointly optimize a multilinear model and the registration of the 3D scans used for training. Given an initial registration, our approach fully automatically improves the registration by optimizing an objective function that measures the compactness of the multilinear model, resulting in a sparse model. We choose a continuous representation for each face shape that allows to use a quasi-Newton method in parameter space for optimization. We show that our approach is computationally significantly more efficient and leads to correspondences of higher quality than existing methods based on linear statistical models. This allows us to evaluate our approach on large standard 3D face databases and in the presence of noisy initializations.

1. Introduction

The human face is one important factor for any kind of social interaction in our daily life. This motivates many different fields such as human computer interaction, medicine, ergonomics or security, to investigate the human face. Since many of these areas are interested in the 3D geometry of the face, the number of publicly available 3D face databases increased over the last years. As the manual analysis of large databases is intractable, automatic data driven and statistical approaches are widely used to analyze the structure of the data. To compute statistics, all shapes of the dataset need to be in correspondence [10, Chapter 1].

Computing these correspondences for human face data is a challenging task that many methods aim to solve (*e.g.* [24, 26, 17, 14, 27]). Given a good registration, a statistical face model can be learned. In computer vision and graphics, statistical face models are used *e.g.* to reconstruct the 3D geometry of the face from 2D images [1], to recognize facial expressions [24], to transfer expressions between images or videos [30], or to change expressions in 3D videos [31].

Statistical face models can also be used to reconstruct the 3D geometry from noisy or partially occluded face scans [5] and are therefore directly applicable for registration. Furthermore, registration methods with prior learned knowledge outperform model-free methods like template fitting [25, 2]. Summing up, this is a chicken-and-egg problem: given a good registration, a statistical model can be learned, and given a representative statistical model, a good registration can be computed. The quality of a given statistical model can be measured [10, Chapter 3.3.1], and due to the dependency of the statistical model on a registration, this measurement also evaluates the underlying registration.

Methods that aim at jointly optimizing the registration and a statistical model have been developed for principal component analysis (PCA) (*e.g.* [10, Chapter 4], [21]). Furthermore, variants of this linear method like part-based PCA [6], kernel PCA [8] or human body specific approaches [16] exist. These methods measure the model quality and change the registration such that the quality of the model and the registration improve at the same time. Since the model quality depends on all shapes, these methods are called groupwise optimization methods. Linear PCA-based methods have been proven to outperform different pairwise correspondence optimization methods [9].

Since the variations in databases of human faces from different identities performing different expressions cannot be modeled well using a linear space, the existing methods are not suitable for optimizing the correspondence of human faces. The space of human faces in various expressions can be well modeled using a multilinear model [30, 24, 31, 2, 5], which is a higher-order generalization of a PCA model.

This motivates us to propose an approach to optimize the correspondence for 3D face databases based on multilinear statistical models. The correspondence is optimized based on the minimum description length (MDL) principle, which leads to a sparse multilinear model. A key advantage of extending MDL to multilinear models is a reduced parameter space, which can be optimized efficiently. The main challenge is that while for the linear case PCA provides an optimal low-dimensional space, the solution for the multilinear case is NP-hard to compute [15]. To find a good basis for face models, we compare different tensor decompositions for their ability to reconstruct unseen face data. Another previously unaddressed challenge related to 3D face data is to allow for manifold boundaries during optimization, which is needed as the face has a mouth and an outer boundary. We solve this issue efficiently by introducing constraints in the optimization framework.

The main contributions of this work are: (1) we introduce the first fully automatic groupwise correspondence optimization approach for multilinearly distributed data, and (2) we show that our approach is computationally significantly more efficient and leads to correspondences of higher quality than existing PCA-based optimization methods.

2. Related work

Template-based facial correspondence computation: Our method is related to methods that aim to compute correspondences between sets of shapes. While many methods exist to establish correspondence for arbitrary classes of shapes, we focus on 3D face registration methods. Given a sparse set of 3D landmarks, Mpiperis *et al*. [24] register 3D faces in various expressions with an elastically deformable face model. Passalis *et al*. [26] fit an annotated face model to the scan by solving a second order differential equation. Huang *et al*. [17] split the face into multiple parts and perform a deformation of each part to fit an input face. Guo *et al*. [14] use a thin-plate spline guided template fitting to register 3D face scans. Pan *et al*. [25] use a sparse deformable model for registration. They learn a dictionary on a set of registered faces and register a new face by restricting the correspondences to be a sparse representation of the learned dictionary. Salazar *et al*. [27] use a blendshape model to fit the expression followed by a template fitting using a non-rigid iterative closest point method to get the facial details.

All of these methods find a good correspondence, but none of them aim at producing a registration that is optimal for statistical modeling. Note that any of these methods can be used to initialize our optimization approach.

Statistical face models: Given a set of 3D shapes in full correspondence, various methods can perform statistical analysis. We focus our discussion on multilinear shape spaces for 3D faces. Vlasic *et al*. [30] use a multilinear model for a database of human faces that decouples facial shape, expression, and viseme to transfer facial performance between 2D videos. Mpiperis *et al*. [24] use a multilinear model for identity and expression recognition. Yang *et al*. [31] reconstruct the 3D face shape from 2D videos and exploit the decoupling of identity and expression variations to modify the identity or expression within the videos. Bolkart and Wuhrer [2] use a multilinear model to register a large database of 3D faces in motion and perform analysis on the resulting registration. Brunton *et al*. [5] learn multiple localized multilinear models and use these to reconstruct models from noisy and occluded face scans.

As these methods use a multilinear model for 3D faces of multiple identities and expressions, they employ the same model as our method. However, none of them aim at optimizing the correspondence using the learned model.

Registration optimization: While some prior works in machine learning explore the idea of jointly learning a model and correspondence information (*e.g*. [4, 29, 18]), our method is most related to methods that aim to jointly optimize the registration of a set of 3D shapes and a learned statistical model. Kotcheff and Taylor [21] propose a groupwise correspondence optimization based on a PCA model that explicitly favors compact models. Davies *et al*. [10, Chapter 4] give an overview of different objective functions for correspondence optimization and motivate an information theoretic objective function minimizing the description length of the data. The basic concept of minimum description length approaches is to minimize the length of a message that is transmitted from a sender to a receiver. They encode the data with a PCA model and alter the correspondence such that the number of bits needed to describe the model and the encoded data is minimal. Davies *et al*. [9] show that MDL outperforms state-of-the-art registration methods for medical datasets. Gollmer *et al*. [13] compare different objective functions. They show that while the determinant of the covariance matrix is easier to optimize, the results are comparable to results produced by MDL.

All these methods model the data with one linear PCA model. In contrast, Burghard *et al*. [6] use a part-based linear model, and Chen *et al*. [8] model the data with a non-linear kernel PCA. Hirshberg *et al*. [16] derive a skeleton based approach specifically for human body shapes to jointly optimize the registration and a statistical model.

None of these methods can model 3D faces with varying identities and expressions. To allow this, we introduce the first groupwise correspondence optimization approach for multilinearly distributed data. Furthermore, while most methods assume the object to be a closed manifold, our approach handles manifolds with multiple boundaries.

3. Multilinear shape space

This section introduces the multilinear model and different tensor decompositions to derive the model. Multilinear models can effectively model statistical variations of faces due to identity and expression as it decouples these two types of shape variation.

3.1. Multilinear model

Given a set of registered and spatially aligned 3D face scans of d_2 identities in d_3 expressions each, every face is represented by a vector $\mathbf{f} = (x_1, y_1, z_1, \cdots, x_n, y_n, z_n)^T$ that consists of n vertices (x_i, y_i, z_i). We center each face by subtracting the mean over all training faces $\bar{\mathbf{f}}$ and arrange the centered faces in a 3-mode tensor $\mathcal{A} \in \mathbb{R}^{3n \times d_2 \times d_3}$, where modes describe the different axes of a tensor. The data are placed within $\mathcal{A}$, such that the vertex coordinates are associated with the first mode, the different identities with the second mode, and the different expressions with the third mode of $\mathcal{A}$. The decomposition of $\mathcal{A}$ into

$$\mathcal{A} = \mathcal{M} \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \tag{1}$$

where $\times_n$ denotes the n-th mode product, results in a tensor $\mathcal{M} \in \mathbb{R}^{3n \times m_2 \times m_3}$ called multilinear model, and orthogonal factor matrices $\mathbf{U}_2 \in \mathbb{R}^{d_2 \times m_2}$ and $\mathbf{U}_3 \in \mathbb{R}^{d_3 \times m_3}$. The n-th mode product $\mathcal{M} \times_n \mathbf{U}_n$ of tensor $\mathcal{M}$ with matrix $\mathbf{U}_n$ replaces each vector $\mathbf{m} \in \mathbb{R}^{m_n}$ aligned with i-th mode by $\mathbf{U}_n \mathbf{m} \in \mathbb{R}^{d_n}$. The multilinear model represents a registered 3D face $\mathbf{f} \in \mathbb{R}^{3n}$ as

$$\mathbf{f} \approx \bar{\mathbf{f}} + \mathcal{M} \times_2 \mathbf{w}_2^T \times_3 \mathbf{w}_3^T, \tag{2}$$

where $\mathbf{w}_2 \in \mathbb{R}^{m_2}$ and $\mathbf{w}_3 \in \mathbb{R}^{m_3}$ are the identity and expression coefficients.

3.2. Tensor decompositions

The decomposition of $\mathcal{A}$ in Equation 1 is called Tucker decomposition. The goal is to find the best Tucker decomposition with a lower-dimensional tensor $\mathcal{M}$ that is as close as possible to $\mathcal{A}$. The quality of the tensor approximation is measured by the norm of the residual. Computing the best Tucker decomposition is NP-hard [15]. Furthermore, in contrast to decomposing a matrix into orthogonal matrices (computed using singular value decomposition (SVD)), Tucker decompositions are not unique. An exact Tucker decomposition can be computed if $m_n = rank(\mathbf{A}_{(n)})$ for all n. Here, $\mathbf{A}_{(n)}$ denotes the matrix unfolding of $\mathcal{A}$ in the direction of n-th mode (all vectors in the direction of the n-th mode form the columns of $\mathbf{A}_{(n)}$). If $m_n < rank(\mathbf{A}_{(n)})$ for at least one n, the decomposition approximates $\mathcal{A}$.

The following describes different methods to compute the Tucker decomposition. Section 5.1 evaluates for each method its ability to reconstruct unseen data when applied for model fitting. We compare the three tensor decompositions described by Kolda and Bader [20], namely: higher order SVD (HOSVD) [22], higher order orthogonal iteration (HOOI) [22], and a Newton-Grassmann optimization approach [12]. All these methods compute a Tucker decomposition for given maximum mode ranks m_2 and m_3.

HOSVD: HOSVD is a higher-order generalization of matrix SVD. To compute the matrices $\mathbf{U}_n$, a matrix SVD is performed as $\mathbf{A}_{(n)} = \mathbf{U}_n \mathbf{S}_n \mathbf{V}_n^T$, where $\mathbf{U}_n \in \mathbb{R}^{d_n \times d_n}$ contains the left singular vectors of $\mathbf{A}_{(n)}$. Truncating columns then reduces the dimensions of identity and expression space. The multilinear model is then computed as $\mathcal{M} = \mathcal{A} \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T$. Even for given m_2 and m_3, the truncated HOSVD does not give an optimal approximation of $\mathcal{A}$.

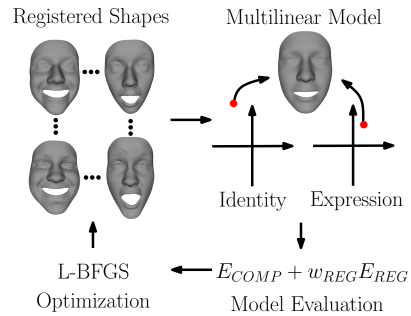

Figure 1. Overview of the iterative multilinear registration.

HOOI: Initialized by HOSVD, this method iteratively optimizes the Tucker decomposition. Within each iteration, both factor matrices are updated by fixing one and updating the other. That is, for a fixed mode-2 factor matrix, a tensor $\mathcal{X} = \mathcal{A} \times_2 \mathbf{U}_2^T$ is computed, and $\mathbf{U}_3$ is updated by the m_3 left singular vectors of $\mathbf{X}_{(3)}$. A similar computation is performed for a fixed mode-3 factor matrix. While HOOI gives a better approximation of $\mathcal{A}$ than HOSVD, it does not necessarily find a stationary point.

Newton-Grassmann optimization: Initialized by HOSVD, the Newton-Grassmann optimization approach constrains each factor matrix to a Grassmannian manifold, an equivalence class of orthogonal matrices. The Tucker decomposition is then computed by a non-linear Newton method on the product of two Grassmannian manifolds. This method converges to a stationary point.

The evaluation of the different tensor decompositions shows that applied to reconstructing unseen face data, they perform almost identical (see Section 5.1). Since HOSVD is the most efficient approach, in the following, we use HOSVD to learn the multilinear model.

4. Groupwise correspondence optimization

This section introduces the concept of groupwise correspondence optimizations and describes our approach for multilinearly distributed data. Given a set of shapes in correspondence, groupwise correspondence optimization minimizes an objective function that measures the quality of the correspondence depending on all shapes. Using a statistical model that describes the variation of the shapes, the objective function measures favorable properties of the model.

For PCA models, Kotcheff and Taylor [21] choose the objective function to be the determinant of the covariance matrix, which explicitly favors the induced linear statistical model to be compact. The compactness of a linear statistical model can be maximized by minimizing the variability of the model, measured by the trace of the covariance matrix.

Compactness measures the variability captured by a model. A compact model can describe instances of a given dataset with the minimum number of parameters and has minimal variance. For models of different compactness that describe the same data, the model with higher compactness and hence lower variance is favorable. It has been shown that minimizing the variance of a PCA model performs similarly to information theoretic approaches that aim at minimizing the description length of the model [13].

Inspired by these previous works, we develop the first MDL-based optimization approach for multilinear models. This extension is challenging because the notion of compactness needs to be extended to multilinear models, where optimal tensor approximation is NP-hard. For 3D face data, a further challenge arises from manifold boundaries. Figure 1 gives an overview of our multilinear optimization approach. Given a set of 3D faces of different identities performing different expressions with an initial correspondence, we iteratively optimize the correspondence. We compute a multilinear model on the registered data, and iteratively improve the model. In each iteration, the quality of the model is measured using a groupwise objective function (Section 4.1). The registered shapes are represented using a continuous parametrization (Section 4.2), and the objective function is optimized in parameter space with a quasi-Newton method (Section 4.3).

4.1. Multilinear objective function

Our groupwise objective function consists of two parts: a compactness energy E_{COMP}, and a regularization energy E_{REG}. We therefore aim to minimize

$$E = E_{COMP} + w_{REG} E_{REG}, \tag{3}$$

where w_{REG} is a weight that controls the influence of the regularization. We now describe both terms in more detail.

Compactness: The compactness of a multilinear model can be measured as the percentage of data variability captured in the first k components of each mode, where $k = 1, \ldots, \max(d_2, d_3)$ [2]. Compactness is maximized by a sparse model that captures all of the variability in few components. To encourage a sparse model, we introduce an energy on the variability of the identity and expression subspaces. Like Kotcheff and Taylor [21], we choose a log-sum penalty function, as log-sum functions are known to encourage sparsity by heavily punishing small values [7]. That is, we aim to minimize

$$E_{COMP} = \frac{1}{d_2}\sum_{i=1}^{d_2} \ln(\lambda_i^{(2)} + \delta_2) + \frac{1}{d_3}\sum_{i=1}^{d_3} \ln(\lambda_i^{(3)} + \delta_3), \tag{4}$$

where $\lambda_i^{(n)}$ denotes the i-th eigenvalue of the mode-n covariance matrix. Small regularization constants δ_n are used

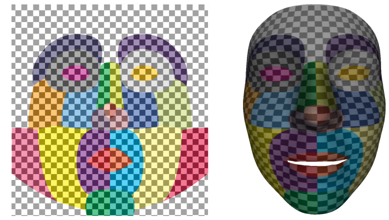

Figure 2. Initial surface parametrization of the 3D face template. Left: 2D parameter domain. Right: 3D parametrization.

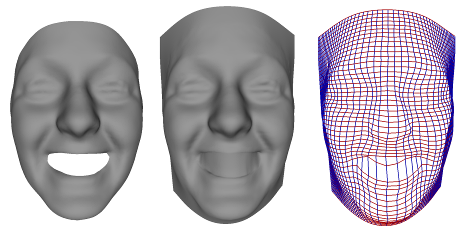

Figure 3. Parametrization for one shape. Left: initialization. Middle: thin-plate spline. Right: (u, v)-parameter lines.

to avoid singularities of E_{COMP} for vanishing eigenvalues. Equivalent to HOSVD, the mode-2 and mode-3 covariance matrices are computed as $\frac{1}{d_3}\mathbf{A}_{(2)}\mathbf{A}_{(2)}^T$ and $\frac{1}{d_2}\mathbf{A}_{(3)}\mathbf{A}_{(3)}^T$.

The energy E_{COMP} is minimized by moving points within the continuous surface of each shape. Since the computation of the covariance only considers a discrete number of points instead of the continuous surface, E_{COMP} can be minimized by moving points away from complex geometric regions with high variability.

Regularization: To avoid undersampling in these regions, Davies *et al.* [10] approximate the integral of the continuous covariance matrix by weighting the points by their surrounding surface area. Since this does not always prevent the undersampling [13], as done in Burghard *et al.* [6], we use a regularization within the objective function. The regularization term for each shape is a bi-Laplacian of the form

$$E_{REG} = \frac{1}{n}\sum_{k=1}^{n} \left\| U^2(\mathbf{v}_k(\mathbf{x})) \right\|^2, \tag{5}$$

where $\mathbf{v}_k(\mathbf{x})$ denotes the k-th vertex of shape $\mathbf{x}$. The double-umbrella operator $U^2(\mathbf{p})$ is the discrete bi-Laplacian approximation [19] computed by

$$U^2(\mathbf{p}) = \frac{1}{|N(\mathbf{p})|}\sum_{\mathbf{p}_r \in N(\mathbf{p})} U(\mathbf{p}_r) - U(\mathbf{p}), \tag{6}$$

where $N(\mathbf{p})$ denotes the set of neighbors of vertex $\mathbf{p}$ within the mesh, and $U(\mathbf{p}) = 1/|N(\mathbf{p})| \sum_{\mathbf{p}_r \in N(\mathbf{p})} \mathbf{p}_r - \mathbf{p}$. The bi-Laplacian regularizer encourages the points to be regularly distributed over the mesh and prevents fold-overs.

4.2. Parametrization

The registration is optimized by moving points in the surface of each face. Since the surface of the face is 2-dimensional, moving points within the surface can be done by re-parametrization. This requires an initial parametrization together with a continuous mapping from parameter space to the surface of each face. We compute an initial registration for a database of 3D faces using template fitting, and additionally unwrap the 3D template mesh in 2D parameter space to compute an initial discrete parametrization with parameters $\mathbf{t}_i \in \mathbb{R}^2$. The embedding in 2D is chosen to minimize distortions of angles and areas. Each parameter $\mathbf{t}_i$ is mapped to the mesh vertex $\mathbf{v}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$. Figure 2 visualizes the initial parametrization in 2D parameter space (left) and mapped on the 3D surface (right). Due to the full correspondence of all face shapes, this discrete parametrization is the same for all shapes of the database.

With this discrete embedding in parameter space, a continuous mapping Φ is computed that maps parameters $\boldsymbol{\alpha} = (u, v) \in \mathbb{R}^2$ into the surface of the shape. A thin-plate spline [11] defines this mapping, computed as

$$\Phi(\boldsymbol{\alpha}) = \mathbf{c} + \mathbf{A}\boldsymbol{\alpha} + \mathbf{W}^T(\sigma(\boldsymbol{\alpha} - \mathbf{t}_1), \ldots, \sigma(\boldsymbol{\alpha} - \mathbf{t}_n))^T, \quad (7)$$

where $\mathbf{c} \in \mathbb{R}^3$, $\mathbf{A} \in \mathbb{R}^{3\times 2}$, and $\mathbf{W} \in \mathbb{R}^{n\times 3}$ are the parameters of the mapping, and where $\sigma : \mathbb{R}^2 \to \mathbb{R}$ is the function

$$\sigma(\mathbf{h}) = \begin{cases} \|\mathbf{h}\|^2 \log(\|\mathbf{h}\|) & \|\mathbf{h}\| > 0, \\ 0 & \|\mathbf{h}\| = 0. \end{cases} \quad (8)$$

The surface of Φ interpolates all vertices of the shape ($\Phi(\mathbf{t}_i) = \mathbf{v}_i$) and gives the surface with the minimum bending energy. Figure 3 shows one initially registered shape (left) together with the computed continuous thin-plate spline visualized as densely approximated mesh (middle) and (u, v)-parameter lines (right). The evaluation of Φ at parameters $\boldsymbol{\alpha}$, where u (respectively v) is fixed and v (respectively u) is varied by a fixed discrete step size, gives one (u, v)-parameter line. While the spline interpolates the geometry of the initial shape, it gives a reasonable extrapolation of the shape beyond the outer border of the face.

4.3. Optimization

The objective function E in Equation 3 is non-linear. Due to the choice of the parametrization, E is analytically differentiable with respect to $\boldsymbol{\alpha}$. The supplementary material gives the full analytical gradient. We minimize E using L-BFGS [23], a quasi-Newton method with linear constraints. These linear constraints allow for each vertex in parameter space to specify a valid rectangular area.

Boundary constraints: For meshes with boundary, E_{COMP} is minimized if the entire surface collapses into a single point. Hence, boundary conditions need to be enforced. Face shapes have two boundaries, an inner boundary at the mouth and an outer boundary at the end of the acquired scan. Since landmarks are used during the initial registration, the inner boundary at the mouth is registered well. To avoid points that move from the lower to the upper lip or vice versa, we fix the points in the 1-ring neighborhood of the mouth boundary during optimization. Since the outer boundary is not registered well as scans in the database are cropped inconsistently, we allow limited movement for points in the 1-ring neighborhood of the outer boundary. Specifically, the movement is restricted to at most 20 mm.

Optimization schedule: Optimizing for the parameters of all shapes at the same time is not feasible for a large population of shapes due to the large number of parameters ($d_2 d_3 2n$). Instead, we only optimize the parameters of each shape individually as proposed by Davies *et al*. [10, Chapter 7.1.1]. This optimization is performed for all shapes of the database during each iteration. Note that E still depends on all shapes for this shape-wise optimization, and the method therefore still optimizes the groupwise correspondence. To avoid bias towards any shape, the order of the shapes is randomly permuted for each iteration step. Since the rigid alignment of the shapes depends on the correspondence, during optimization of one shape, the alignment is updated after a few optimization steps.

Computational complexity: The computational complexity of one optimization step is $O(nd_2^2 d_3 + nd_2 d_3^2)$ (see supplementary material for details). As shown in the following section, our approach is significantly more efficient than existing PCA-based MDL approaches.

5. Evaluation

This section evaluates three different tensor decompositions and our model optimization approach.

Data: For evaluation, we use models of the BU-3DFE [32] and Bosphorus [28] databases. BU-3DFE contains 3D face scans in neutral expression and in six prototypic expressions. Bosphorus covers the six prototypic expressions and a subset of up to 28 action units per subject. Since both databases are acquired with different scanner systems, the resulting scans have different resolution and noise characteristics. We register the face scans with a template fitting method [27] using the provided landmarks.

For BU-3DFE we use 50 randomly chosen identities in 7 expressions: neutral and the highest level of each expression. For Bosphorus we use all 65 identities that are present in all 7 expressions. In the following, we call these subsets BU-3DFE subset and Bosphorus subset, respectively.

Model quality: We quantitatively evaluate the quality of the optimization with the widely used measures compactness, generalization and specificity [10, Chapter 9.2]. The

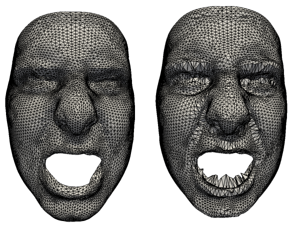

Figure 4. Artifacts obtained by optimizing E_{COMP} without regularization ($w_{REG} = 0$). Left: initial registration. Right: result.

identity and expression spaces should ideally be compact, general and specific.

Generalization measures the ability of the statistical model to represent shapes that are not part of the training. The generalization error is measured in a leave-one-out fashion. For the identity mode, each subject is once fully excluded from training and the resulting model is used to reconstruct all excluded scans. The error is then measured as the average vertex distance between all corresponding vertices. The error for the expression mode is computed accordingly by excluding once each expression.

Specificity measures the ability of the statistical model to only represent valid shapes of the object class. To measure the specificity of the model before and after optimization, we randomly choose 10000 samples in identity and expression space and measure the average vertex distance of the reconstruction to the training data.

Reproducibility: To facilitate evaluating the model for different applications, we make our optimization code and the optimized statistical model available [3].

5.1. Tensor decompositions

We evaluate the different tensor decomposition methods described in Section 3.2 by fitting the resulting multilinear models to unseen 3D face scans. For this, we use a 10-fold cross validation on the registered BU-3DFE scans. We split the database randomly into ten groups, each with the same ratio of male and female subjects, where all scans of one identity belong to the same group. The error is measured as the distance between a vertex in the fitting result and its closest point in the face scan. The error distribution of all three methods is nearly identical. The median vertex error is for HOSVD 1.145 mm, for HOOI 1.144 mm and for the Newton Grassmann method 1.144 mm. Since all methods perform almost the same, we compute the decomposition with HOSVD in the following.

5.2. Influence of regularization

This section evaluates the influence of the regularization E_{REG} on the BU-3DFE subset. The optimization is performed twice, once only optimizing E_{COMP} without E_{REG} and once only optimizing E_{REG} without E_{COMP}. As discussed in Section 4.1, the regularizer is needed to

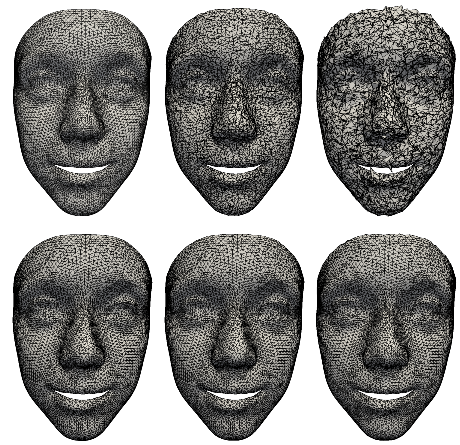

Figure 5. Noise example of the database before (top) and after (bottom) optimization. Left to right: no, low, and high noise.

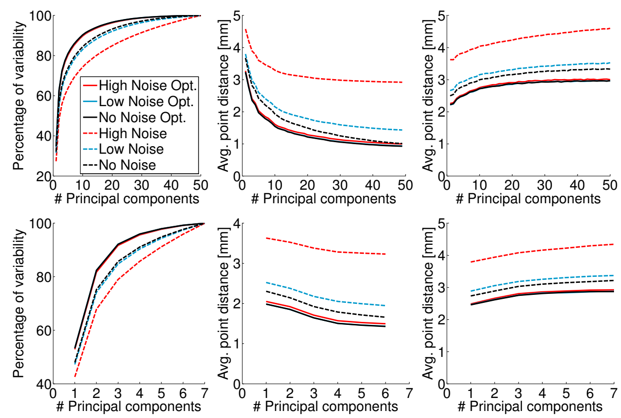

Figure 6. Influence of the initialization for different levels of noise. Left: compactness. Middle: generalization. Right: specificity. Top: identity mode. Bottom: expression mode.

avoid undersampling in regions with high variability and fold-overs. Figure 4 shows the result for one face after only five iterations of optimizing E_{COMP}. When minimizing only E_{COMP}, the optimization moves points away from the eyebrows and around the nose, resulting in sparsely sampled regions. Furthermore, fold-overs at the mouth cause visual artifacts. Optimizing E_{REG} leads to regularly sampled meshes. However, E_{COMP} increases in this case. Minimizing E is therefore a tradeoff between getting a compact model and a regular mesh structure. In the following, we empirically choose $w_{REG} = 0.5$.

5.3. Influence of initialization

This section evaluates the robustness to noise in the initialization. State-of-the-art registration methods for faces, as used for the initialization of our method, are able to fit the facial surface well with sub-millimeter accuracy, but the result is likely to contain drift within the surface. To simulate noise regarding these methods, we use the initial

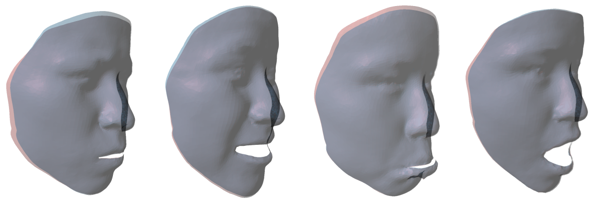

Figure 7. Visual comparison of template fitting [27] (red) and our result (blue) for one subject in four expressions (overlap in gray).

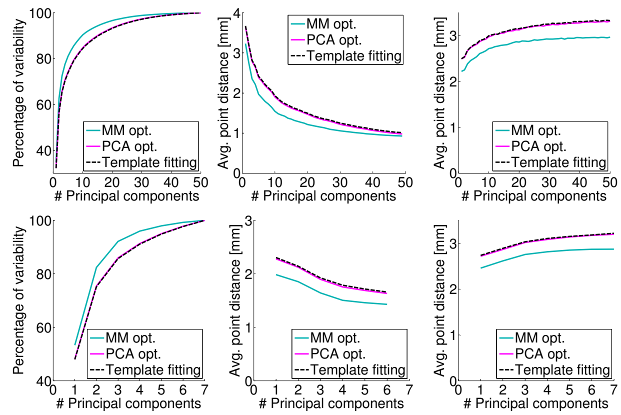

Figure 8. Comparison of template fitting [27], PCA optimization [10] (PCA opt.) and multilinear model optimization (MM opt.) on BU-3DFE subset. Left: compactness. Middle: generalization. Right: specificity. Top: identity mode. Bottom: expression mode.

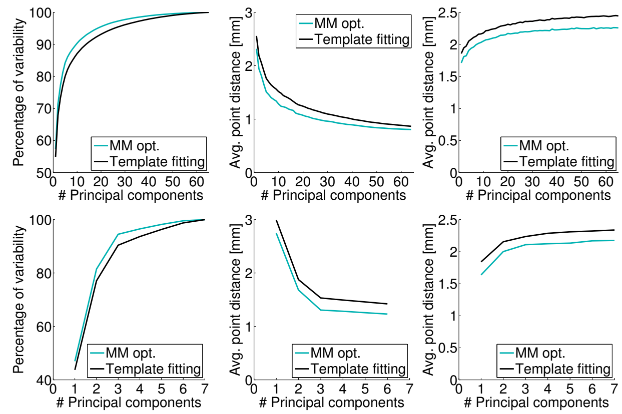

Figure 9. Comparison of template fitting [27] and multilinear model optimization (MM opt.) on Bosphorus subset. Left: compactness. Middle: generalization. Right: specificity. Top: identity mode. Bottom: expression mode.

parametrization and add two different levels of noise in the parameter domain. The parameter values of each shape of the BU-3DFE subset are disturbed by random Gaussian noise. Since the 1-ring neighborhood of the mouth boundary is fixed during optimization, these vertices are left without noise. For both noise levels we choose noise with mean zero and standard deviation f times the average 3D edge length. For the lower noise level we choose f to be 0.25, for the higher 0.75, respectively.

The optimization is performed on the BU-3DFE subset, initialized with the noisy registration. The top of Figure 5 shows an example of the database without noise (left), the lower level of noise (middle) and the higher level of noise (right). The average 3D vertex distance of the initial shapes to the noisy shapes over the entire database is 1.11 mm for the lower and 2.50 mm for the higher noise level.

Adding random noise within the surface to each vertex increases the variance in 3D positions and therefore increases the variability of the data. As expected, Figure 6 shows that the compactness of identity mode and expression mode decreases with increasing noise, since the multilinear model captures less variability with the same number of components. Further, the multilinear model becomes less general and less specific. After 15 iterations, the average compactness increases by 3.8% for the low noise level, and by 8.7% for the high noise level, respectively. The average generalization error decreases by $0.58mm$ and $1.65mm$ for the low and high noise level, the average specificity decreases by $0.43mm$ and $1.26mm$ for the low and high noise level. After optimization, the model quality for both levels of noise is comparable to the optimization of the data without noise. Hence, our optimization method effectively reduces variability caused by drift.

5.4. Comparison

This section compares our approach to two state-of-the-art registration methods for 3D faces based on template fitting [27] and PCA-based groupwise correspondence [10].

Template fitting: We compare our optimization to template fitting on the BU-3DFE and Bosphorus subsets. For the two subsets, Figures 8 and 9 show the compactness, generalization and specificity for template fitting and after 15 iterations of the multilinear optimization. For the BU-3DFE subset, the average compactness increases by 3.0%, and the average generalization and specificity decrease by $0.25mm$ and $0.32mm$, respectively. For the Bosphorus subset, the average compactness increases by 1.7%, and the average generalization and specificity decrease by $0.15mm$ and $0.16mm$, respectively.

Figure 7 visually compares the template fitting (red) to our result (blue) for one subject of the BU-3DFE subset. Before optimization, the shape of the outer boundary differs. The optimization decreases the face for the first and fourth expressions at the cheek, for the second expression at the jaw, and for the third expression at the forehead. Expressions one, two and three are extended at the forehead. After 15 iterations, the outer boundaries are similar.

To demonstrate the ability of our method to optimize over large sets of shapes, we consider a second subset of the Bosphorus database consisting of 39 identities performing

26 action units each, leading to a total of over 1000 shapes. To keep 95% of the data variability after template fitting, a total of 27 components are necessary, while after 15 iterations of our optimization, 20 components suffice. As for the other subsets, generalization and specificity also improve after optimization. To the best of our knowledge, this is the first time a registration optimization based on MDL has been applied to such a large set of shapes.

For all three datasets the model improves significantly during optimization, leading to a more compact model with improved generalization and specificity.

PCA: For brevity, we abbreviate PCA optimization by PCA opt. and our method by MM opt. during the discussion of the comparison. We start by comparing the computational complexity of the two methods. In the supplementary material, we show that one optimization step for PCA opt. has complexity $O(nd_2^2d_3^2)$, while one optimization step of MM opt. has complexity $O(nd_2^2d_3 + nd_2d_3^2)$. For the BU-3DFE subset our non-optimized implementation takes about 16.2h for MM opt. and about 21.5h for PCA opt. for one iteration when executed on a standard PC.

Figure 8 quantitatively compares PCA opt. and MM opt., both after 15 iterations. While MM opt. gives significant improvements, PCA opt. only slightly improves the correspondence. For small subsets PCA opt. gives significant improvements within few iterations. Our experiments suggest that for an increasing number of shape space parameters, an increasing number of iterations is required. Since MM opt. models identity and expression independently, the number of shape space parameters is $d_2 + d_3$, while for PCA opt. the number of shape space parameters is d_2d_3.

Hence, our method gives better improvements after the same number of iterations and is computationally faster than existing linear optimization methods.

5.5. Discussion

Parametrization: Our proposed method optimizes the correspondence by re-parametrizing the shapes guided by the optimization of a multilinear compactness objective function. This re-parametrization requires a continuous representation of the surface for each shape. While any kind of continuous mapping can be used, we establish this by a thin-plate spline. For other continuous mappings, the gradient changes, and therefore depending on the mapping (*e.g*. for mappings without analytical gradient) E must be optimized with a different method.

Data quality: Computing this continuous surface mapping assumes the original face scans to be regularly densely sampled with points that are within the surface of the scan. To get this sampling, any existing template fitting method can be used. For face scans with partial occlusions or strong distortions, template fitting methods fail, since they are unable to estimate the real face surface in these regions. To optimize the registration for scans with strong distortions, we would either need another initialization that gives a reasonable surface estimation within the occluded and noisy regions (*e.g*. Brunton *et al*. [5]), or the optimization of E must be allowed to leave the surface of the disturbed scan guided by the underlying multilinear model.

Computational complexity: While the multilinear correspondence optimization is computationally more efficient than previous linear methods, due to the groupwise objective function, the computational complexity is still high. Our experiments show that only a low number of iterations are necessary to get significant improvements. Note that the registration can be seen as pre-processing that only needs to be done once. The application for larger datasets would require the use of a compute cluster to exploit the full potential of the parallelizability of the method (especially the gradient computation).

Extensions: Our method is generally applicable to other classes of multilinearly distributed data. The geometry of the shapes can contain no or multiple holes as long as the boundaries of the holes are constrained. The regularization E_{REG} prevents fold-overs around these holes. Furthermore, the extension of our method to more modes is straightforward, *e.g*. for faces to associate the fourth mode with viseme or age.

6. Conclusion

We have presented the first method for multilinearly distributed data that jointly improves a given registration and a multilinear model. A continuous representation of each shape allows to optimize the registration with a quasi-Newton method. We have evaluated our method on scans of two databases and have demonstrated that our method is robust to noise in the initial registration. A key advantage of our approach over existing linear MDL methods is its increased computational efficiency, which allows for the first time to apply an approach based on MDL to databases containing over 1000 shapes. We have shown that using the efficient HOSVD method to compute the multilinear model performs similarly when reconstructing unseen face data to more elaborate tensor decompositions. To facilitate experiments for different application scenarios, we make our optimization code and the optimized statistical model available.

Acknowledgments

We thank Arnur Nigmetov for help with the comparison of the different tensor decompositions, and Alan Brunton, and Michael Wand for helpful discussions. This work has been partially funded by the German Research Foundation (WU 786/1-1, Cluster of Excellence MMCI).

References

[1] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *SIGGRAPH*, pages 187–194, 1999.

[2] T. Bolkart and S. Wuhrer. 3D faces in motion: Fully automatic registration and statistical analysis. *CVIU*, 131:100–115, 2015.

[3] T. Bolkart and S. Wuhrer. Multilinear mdl for 3D faces, 2015. http://multilinear-mdl.gforge.inria.fr/.

[4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *CVPR*, volume 2, pages 690–696, 2000.

[5] A. Brunton, T. Bolkart, and S. Wuhrer. Multilinear wavelets: A statistical shape space for human faces. In *ECCV*, pages 297–312, 2014.

[6] O. Burghard, A. Berner, M. Wand, N. J. Mitra, H. Seidel, and R. Klein. Compact part-based shape spaces for dense correspondences. *CoRR*, abs/1311.7535, 2013.

[7] E. Candès, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted l_1 minimization. *JFAA*, 14(5-6):877–905, 2008.

[8] J.-H. Chen, K. C. Zheng, and L. G. Shapiro. 3D point correspondence by minimum description length in feature space. In *ECCV*, pages 621–634, 2010.

[9] R. Davies, C. Twining, T. Cootes, and C. Taylor. Building 3-d statistical shape models by direct optimization. *MI*, 29(4):961–981, 2010.

[10] R. Davies, C. Twining, and C. Taylor. *Statistical Models of Shape: Optimisation and Evaluation*. Springer, 2008.

[11] I. Dryden and K. Mardia. *Statistical shape analysis*. Wiley, 1998.

[12] L. Eldén and B. Savas. A newton-grassmann method for computing the best multilinear rank-(r_1, r_2, r_3) approximation of a tensor. *SIAM J. Matrix Anal. Appl.*, 31(2):248–271, 2009.

[13] S. T. Gollmer, M. Kirschner, T. M. Buzug, and S. Wesarg. Using image segmentation for evaluating 3D statistical shape models built with groupwise correspondence optimization. *CVIU*, 125(0):283 – 303, 2014.

[14] J. Guo, X. Mei, and K. Tang. Automatic landmark annotation and dense correspondence registration for 3D human facial images. *BMC Bioinf.*, 14(1), 2013.

[15] C. J. Hillar and L.-H. Lim. Most tensor problems are NP-hard. *JACM*, 60(6):45:1–45:39, 2013.

[16] D. Hirshberg, M. Loper, E. Rachlin, and M. Black. Coregistration: Simultaneous alignment and modeling of articulated 3D shape. In *ECCV*, pages 242–255, 2012.

[17] Y. Huang, X. Zhang, Y. Fan, L. Yin, L. Seversky, J. Allen, T. Lei, and W. Dong. Reshaping 3D facial scans for facial appearance modeling and 3D facial expression analysis. *IVC*, 30(10):750 – 761, 2012.

[18] M. Irani. Multi-frame correspondence estimation using subspace constraints. *Int. J. Comput. Vision*, 48(3):173–194, 2002.

[19] L. Kobbelt, S. Campagna, J. Vorsatz, and H.-P. Seidel. Interactive multi-resolution modeling on arbitrary meshes. In *SIGGRAPH*, pages 105–114, 1998.

[20] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, 2009.

[21] A. C. Kotcheff and C. J. Taylor. Automatic construction of eigenshape models by direct optimization. *Med. Image Anal.*, 2(4):303 – 314, 1998.

[22] L. D. Lathauwer. *Signal processing based on multilinear algebra*. PhD thesis, K.U. Leuven, Belgium, 1997.

[23] D. Liu and J. Nocedal. On the limited memory method for large scale optimization. *Math. Prog.: Series A and B*, 45(3):503–528, 1989.

[24] I. Mpiperis, S. Malassiotis, and M. G. Strintzis. Bilinear models for 3-D face and facial expression recognition. *IFS*, 3:498–511, 2008.

[25] G. Pan, X. Zhang, Y. Wang, Z. Hu, X. Zheng, and Z. Wu. Establishing point correspondence of 3D faces via sparse facial deformable model. *IP*, 22(11):4170–4181, 2013.

[26] G. Passalis, P. Perakis, T. Theoharis, and I. Kakadiaris. Using facial symmetry to handle pose variations in real-world 3D face recognition. *PAMI*, 33(10):1938–1951, 2011.

[27] A. Salazar, S. Wuhrer, C. Shu, and F. Prieto. Fully automatic expression-invariant face correspondence. *MVAP*, pages 1–21, 2013.

[28] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3D face analysis. In *BIOID*, pages 47–56, 2008.

[29] L. Torresani, D. Yang, E. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *CVPR*, pages 493–500, 2001.

[30] D. Vlasic, M. Brand, H. Pfister, and J. Popović. Face transfer with multilinear models. *SIGGRAPH*, 24(3):426–433, 2005.

[31] F. Yang, L. Bourdev, J. Wang, E. Shechtman, and D. Metaxas. Facial expression editing in video using a temporally-smooth factorization. In *CVPR*, pages 861–868, 2012.

[32] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato. A 3D facial expression database for facial behavior research. In *FG*, pages 211–216, 2006.